International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT217334

Deep fake : An Understanding of Fake Images and Videos

Shweta Negi[*], Mydhili Jayachandran, Shikha Upadhyay
Department of Forensic Sciences, Jain (Deemed-to-be University), Bengaluru, Karnataka, India

ABSTRACT

Article Info
Volume 7, Issue 3
Page Number: 183-189

Publication Issue :
May-June-2021

Article History
Accepted : 10 May 2021
Published : 17 May 2021

The Deepfake algorithm allows its user to create fake images, audios, videos that gives very real impression but is fake in real sense. This degree of technology is achieved due to advancements in Deep Learning, Machine Learning, Artificial Intelligence and Neural Networking that is a combination of algorithms like generative adversarial network (GAN), autoencoders etc. Any technology has its positive and negative repercussions. Deep fake can come in use for helping people who have lost their speech to give them new improved voice, commercially deepfake can be used in improving animation or movie quality putting in creative imagination to work as well is therapeutic to people who have lost their dear once. Negative aspects of deep fake include creating fake images, videos, audios that look very real can cause threats to an individual's privacy, organizations, democracy, and even national security. This review paper presents history on how deep fake emerged, will comprehend on how it works including various algorithms, major research works done on understanding deep fakes in the literature and most importantly discuss recent advancements in detection of deep fake methods and its robust preventive measures.

Keywords: Deepfake, GAN, Autoencoders, Fake images, Fake videos, Artificial Intelligence, Neural Network.

I. INTRODUCTION

The term "deepfake" emerged in late 2017 when a redditor (someone who uses reddit platform which is an American social media platform for web content rating by votes along with discussion of websites) posted realistic pornographic videos featuring Hollywood actresses who weren't really part of it, the user's handle read "deepfake" were the name deep fake came from.[1]

Deep fake requires huge data to train models or neural networks to create fake image, video, audios. A neural network is combination of many neuron which is multilayered. A neuron is a thing that holds a number between 0-1.

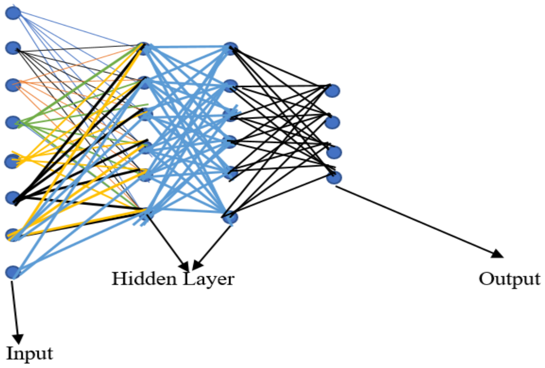
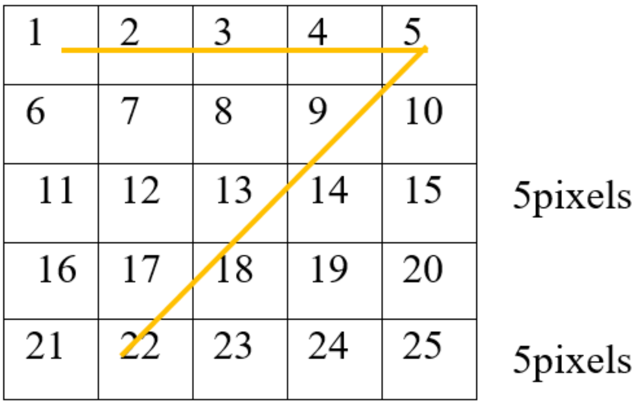


Figure 1.1 Multi layered neuron network

In fig.1.1 different neural layers or multilayer is depicted (Author's creation), 1st layer is the layer that takes in input, the in between two layers is known as hidden layer and the last layer is known as output layer. Here, in diagram 1.2 (Author's creation from different resources) there is total of 5 X 5= 25 pixels in total. Each Pixel represent or remembers part of a figure no.7 (Author's creation from different resources)

For example.

Here in fig. 1.2 total no. of pixels is 5 X 5 = 25 Each pixel contains a specific pattern, here in diagram 2 no. 1.2 is represented by 25 pixels, each pixel has to remember a particular patter of no.7



1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

Figure 2.2

As there are pixels similarly there are neurons that hold the same value as a pixel and are in same quantity for e.g. For fig. no. 1.3 the no. of neurons will be 25 similar to no. of pixels as well as it will hold the same value.

For example, if we look at the first row itself each pixel contains parts of no. 7 as shown in fig. 1.3. pixel contains parts of no. 7 as shown in fig. 1.3.

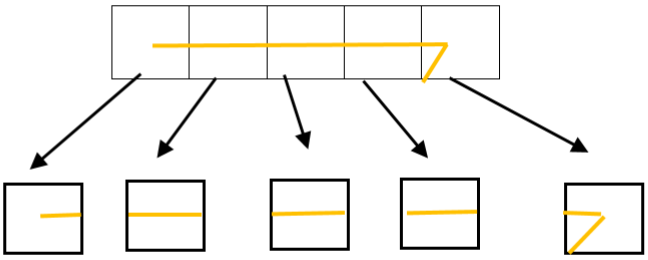
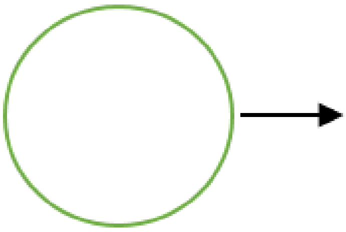


Figure 3.3

As there are pixels similarly there are neurons that hold the same value as a pixel and are in same quantity for e.g. For fig. no. 1.3 the no. of neurons will be 25 similar to no. of pixels as well as it will hold the same value.



It is neuron, a neuron is a thing that holds number and the no. it holds is known as its activation and the value of activation is between 0-1.

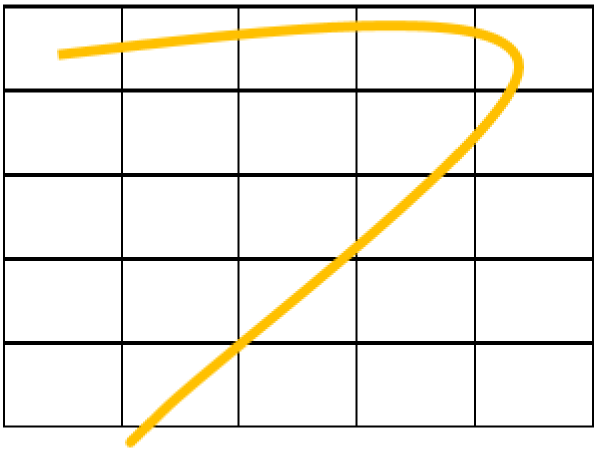


Figure 1.4

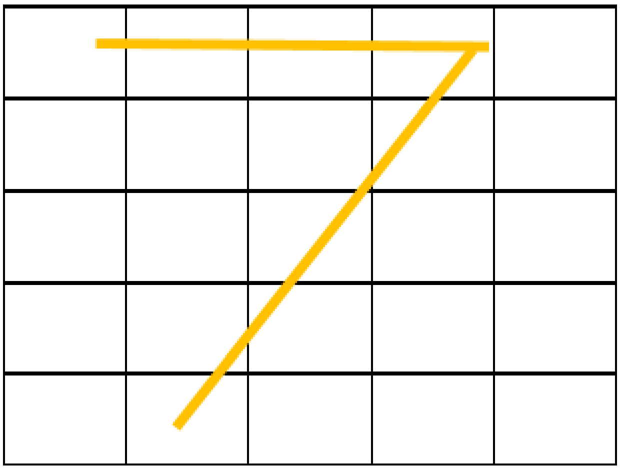


Figure 1.5

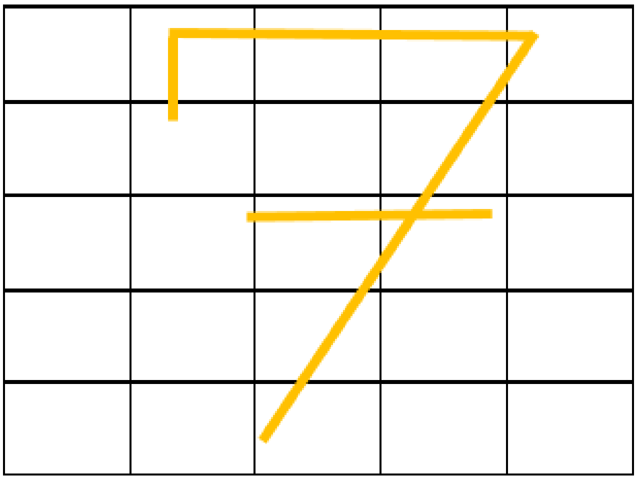


Figure 1.6

In above diagram no. 1.4, 1.5, 1.6 [All depicts no. seven but are written in different manner.]

Manner in which a computer identifies all these as seven itself but not any other numbers.

In fig. no. 1.1 the first layer i.e., input layer will be made up of all 25 neuron that represents no. seven.

The second and third layer i.e. hidden layer will contain different activation and there is connection of various neurons between all layers, the input layer will match those parts of hidden layer neuron that matches the symbol seven, and confirms the output with output layer. If the output layer tells that it is wrong the neural network tweak itself until correct output is provided and that is how it self learns and starts recognizing various patterns, post that a huge data set is required to train neural networks before actually implementing it on field. As activation value is between 0 to 1 a neural network will recognize that two things are similar if the activation value is above 0.5 and vice versa.

This was a very basic idea of neural network and how it works, neural networks in deep fakes are different and more complex. Generative adversarial networks i.e., GANs are responsible to create those fake videos or images of faces that don't exist, and doctor those images and videos. These face generators are made up of basic network known as neural network.

A neural network takes input does some processing and gives output. If problem for processing is face generation its more complex because the network has to reads in input and then extract features like eyes, nose, mouth, texture, facial features, determine contort of such features and much more, neural network to generate face should have high processing and a large data and a lot of time as well as complex neural network to understand and train in face recognition.[4]

II. METHODOLOGY

Methods of creating and detecting deepfake

a) Creation- Method of creating deepfake includes involvement of encoder decoder network as well as (CNN), (RNN), combination of their techniques etc. [3,7,11,14]

b) Detection-Method of detecting deepfake include separate parameters of both image and video. Fake image detection involves detection of gaussian noise, blur and fake video involves temporal features across video frames. [11,13,14]

III. RESULTS AND DISCUSSION

Metamorphosis of Neural Networks.

Understanding about Neural Networks is since 1943, but the hardware never had enough processing power until recently. Processing power of computer grew exponentially in last few decade. 2010-2012 neural network boomed, some of neural network's applications are language translation e.g. Google translator, in security and defense systems, image captioning Facebook tagging, object localization, object detection, media and entertainment, image and video segmentation, autonomous car, speech recognition, medicine and biology, face generation. Neural networks that do face generation is called generative adversarial network (GANs).[4]

These GANs are made up of many simple networks that are multilayered to form complex networks.

1. GANs (Generative adversarial network)

There are two neural networks i.e., neural network1 and neural network2

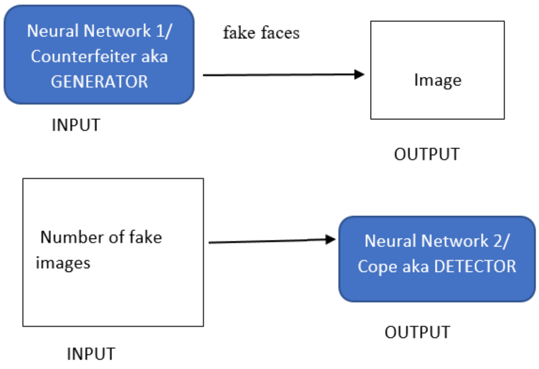

Neural Network 1/ Counterfeiter aka GENERATOR
fake faces
Image
INPUT
OUTPUT
Number of fake images
Neural Network 2/ Cope aka DETECTOR
OUTPUT
INPUT

Figure 4.7 Flow chart of deep fake process

Neural network 2 looks at some fake image and should try to tell which are fake and real images. Neural network 1 and 2 play game where they take turns, its round 1. Counterfeiter generates a fake image and it puts this image in pile that contain two type of images both fake and real ones now its cops turn it takes first image from the pile and answers the question Is this image real or fake? If cope answers questions correctly than he wins otherwise the counterfeiter wins and after each round the looser tweaks itself to improve its performance, so after many such round cop becomes better and better at detecting fakes and counterfeiter becomes better and better at generating fakes and at end of the round, we will ask counterfeiter to generate realistically looking fake image.[4]

GANs are not that efficient and the image quality is poor, but in 2015 researchers Luke Metz and Alec Radford suggested that instead of using simple networks to use complex networks, we can use complex networks to form even more complex networks. Simple multilayered perceptron became complex convolutional networks.

2. Complex Convolutional Networks / Deep Convolutional GANs (DCGAN)

These networks showed more processing results. A DCGAN working is very similar to GANs, but uses Deep Convolutional networks (these algorithms give importance on learnable weights and biases and compare them) [6] rather than fully-connected networks. Convnets in general find areas of correlation within an image, that is, they look for spatial correlations. This means a DCGAN is suitable for image/video data, whereas GAN can be applied to wider domains.[5] Use of complex convolutional networks showed more promising results. Around the same time a new GAN was introduced called Coupled GAN or Co-GAN.[4]

3. Coupled GAN or Co-GAN

Instead of using one generator or one detector it uses two generators and two detectors, hence there will be two simultaneous games played in each round. Each generator networks share information with each other but also slightly tweak themselves to fool corresponding detectors, the end result is that the two generators form two slightly different images e.g. A person with blond hair and same person with brown hair

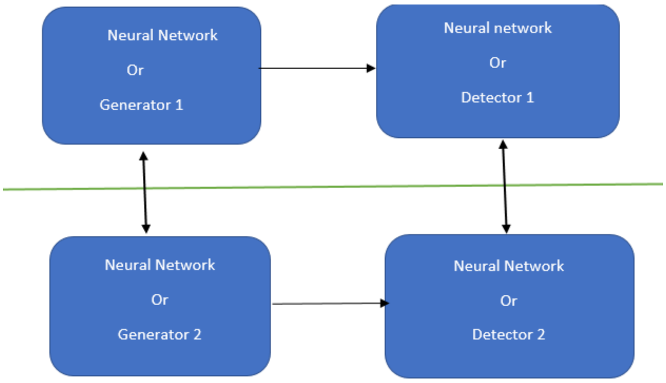


Figure 5.8 Generator and Detector Network

Over time we have seen different types of GANs learning to generate faces but all share the same problem i.e., the generative image is not of high quality as a detector would always tell the image is fake if generator always created high quality or resolution images hence the generator makes sure that image quality is low which is a disadvantage as we get poor quality image from GANs, but this changed after 2018 when NVIDEA was introduced.[4]

Nvidea

The generator and detector play many rounds using similar images of similar quality throughout, but we start generator and detector as simple networks eg. 100 rounds, because of simple network a generator will have to only generate low quality images and the detector won't be effective in telling difference between real and fake images. After 100 rounds we make both networks slightly more complex by adding an additional layer and using high resolution images progressively as rounds go on generator will generate high quality images. At last, we get images that are difficult for humans to even distinguish

But researchers didn't end their work they wanted control over the images being generated which means if we give input as brown and smiling face the network should give the same output this method uses slightly different generative and more complex network.[4]

IV. RESULT

The result for all detection system showed the detection error. The results showed that lip-syncing based algorithm was not able to detect face swapping, as GANs are able to generate facial expressions with high quality that can match audio speech [10,14] which implies that only image based has high accuracy to detect deepfake videos, in future more advanced techniques for face swapping will be difficult to detect. The table explaining methodology as follows.

SR NO.	METHOD	CREATION
1.	**Encoder Decoder Network**	For creating deep fake using encoder networks CNN, RNN, Mixed Style, FC etc and examples for encoder-decoder pairs are Conv Net-RNN, RNN-RNN and LSTM-LSTM
2.	**(a)Fake image detection**	Gaussian blur and Gaussian noise, to remove low level high frequency clues of GAN images
	(b)Fake video detection	Temporal Features across Video Frames

TABLE I: Techniques in deep fake

Some of the prevention methods that were suggested commonly across papers were reliable screening or filtering mechanism throughout all platforms like Social Media, YouTube, etc to automate the ease of detection and remediating spread of fake news, strengthening the legislator and legal requirements, use of watermarking tools into device so that digital content will create immutable metadata that contains information like time, location etc. [3,7,11,14]

Overall subgroup accuracy to detect Deepfake may vary, it is shocking to know that there is statistically little to no major difference between deepfake and original image or video. [11]

V. CONCLUSION

Recent advancement in the field has not proved as effective and the field needs more research to bridge the gap. It was observed that The screening or filtering mechanism using effective detection methods can be implemented on these platforms to ease the deepfakes detection, Legal requirements can be made for tech companies who own these platforms to remove deepfakes quickly to reduce its impacts, In addition, watermarking tools can also be integrated into devices that people use to make digital contents to create immutable metadata for storing originality details such as time and location of multimedia contents as well as their untampered attestment.

VI. REFERENCES

[1]. Clark Merrefield et al (June 27, 2019). Deepfake technology is changing fast - use these 5 resources to keep up. Journalist's Resource.

[2]. J.M. Porup (April 10,2019). How and why deepfake videos work - and what is at risk. CSO India.

[3]. Nguyen, T. T. et al (2019). Deep learning for deepfakes creation and detection. arXiv preprint arXiv:1909.11573, 1.

[4]. CodeEmporium (Jan 19,2019). Evolution of Face Generation | Evolution of GANs [Video]. YouTube.

[5]. Jonathan Hui (June 18,2018). GAN - DCGAN (Deep convolutional generative adversarial networks). Jonathan Hui. [Blog]

[6]. Felix Mohar (Nov 14,2017). Implementing a Generative Adversarial Network (GAN/DCGAN) to Draw Human Faces. towards data sciences.

[7]. Kashif Ali Siddiqui (April 28,2020). Kashif Ali Siddiqui's Answer to "Should we be worried about deep fakes and the misuse of facial recognition?". Quora.

[8]. Buzz Blog Box (Feb 1,2020). How Deepfake Technology Impact the People in Our Society? [Blog].

[9]. Tolosana, R. et al (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion, 64, 131-148.

[10]. Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? assessment and detection. arXiv preprint arXiv:1812.08685.

[11]. Barari, S., Lucas, C., & Munger, K. (2021). Political Deepfake Videos Misinform the Public, But No More than Other Fake Media.

[12]. Guarnera, L., Giudice, O., & Battiato, S. (2020). Deepfake detection by analyzing convolutional traces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 666-667).

[13]. Li, Y.et al (2020). Celeb-df: A large-scale challenging dataset for deepfake forensics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3207-3216).

[14]. Koopman, M., Rodriguez, A. M., & Geradts, Z. (2018, August). Detection of deepfake video manipulation. In The 20th Irish machine vision and image processing conference (IMVIP) (pp. 133-136).

Cite this article as :

Shweta Negi, Mydhili Jayachandran, Shikha Upadhyay, "Deep fake : An Understanding of Fake Images and Videos", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 3, pp. 183-189, May-June 2021. Available at
doi : https://doi.org/10.32628/CSEIT217334
Journal URL : https://ijsrcseit.com/CSEIT217334